# Historical Oral Poems and Digital Humanities : Starting with a Finnish Corpus

Kallio, Kati

2020

Kallio , K , Mäkelä , E & Janicki , M M 2020 , ' Historical Oral Poems and Digital Humanities : Starting with a Finnish Corpus ' , FF Network , no. 54 , 2 , pp. 12-18 . < https://www.folklorefellows.fi/historical-oral-poems-and-digital-humanities/ >

http://hdl.handle.net/10138/330175

unspecified
publishedVersion

# Historical Oral Poems and Digital Humanities

## Starting with a Finnish Corpus

Kati Kallio
Finnish Literature Society and University of Helsinki

Eetu Mäkelä
University of Helsinki Centre for Digital Humanities

Maciej Janicki
University of Helsinki Centre for Digital Humanities

In this essay, we describe early experiments in a computational folkloristics project FILTER[1] aimed at studying formulaic intertextuality, thematic networks and poetic variation across regional cultures of Finnic oral poetry. Due to the vast amount of linguistic and poetic variation and historical biases in the corpora (see e.g. Anttonen 2005; Harvilahti 2013; Tarkka et al. 2018; Ilyefalvi 2018; Mäkelä et al. 2020b), existing automated approaches (see e.g. Moretti 2013) are unusable. Instead, advances must be made through intelligently interleaving computational and manual analysis (Säily et al. 2018; Hämäläinen et al. 2018; Isoaho et al. 2020).

In this project, the idea is to gradually develop tools in tight collaboration between folklorists and computer scientists (Mäkelä et al. 2019; 2020a). The folklorists describe what they tend to do and what they dream of being able to do with the source material, while computer scientists think of what may be possible and how this might be achieved. We first discuss the ideas, proceed to some test computations and then interpret these – and the possible problems – in relation to our humanistic and computational background knowledge of the data itself. If the results seem promising, some prototype interface may be developed, and the folklorists begin experimenting with it, evaluating what does or does not work, and describing what they do so that the computational scientists are able to understand the humanistic needs and the interpretive problems in the data. Folklorists continue dreaming what they would like to do, potentially leading again to new computational solutions and new evaluations in the cycle. In such experiments, even those that are only briefly tried often reveal new aspects of the data and help us to understand it better.

While we aim to build tools and processes that serve our specific project, we are also making them as broadly applicable as possible for researchers working with the same corpus or with similar questions with other materials, particularly for other small languages and oral-derived corpora. On the side of folkloristics, the project builds on the long research history of Finnic oral poems, on advances in computational folkloristics (see e.g. Abello et al. 2012; Arvidson et al. 2018; Harvilahti 2019; Hakamies et al. 2019; Sarv 2019; Tangherlini 2013; 2016) and on discussions with colleagues, especially Frog, Lauri Harvilahti, Janika Oras, Jukka Saarinen, Venla Sykäri and Senni Timonen.

In this essay, we describe our early experiments thus far. At this stage, the main computational question has been how to help the humanist researcher to find relevant sub-corpora or sets of texts, how to tackle complex textual variation, and what tools might be used to find similar, yet varying instantiations of verses and motifs. The central questions have been: (*a*) how to define folkloristically relevant research questions that are narrow enough for the development of new tools and yet help to produce and test tools with potential for wider use; and (*b*) how to analyse and explain the quite complex and versatile processes of reading, contextualising and analysis that folklorists tend to do with historical poetic texts, so that the computational scholars can help to make these processes easier.

### Finnic Oral Poetry and the SKVR Corpus

Historical Finnic oral poetry – *runo*-songs, *regilaul*, or Kalevalaic poetry – makes a versatile corpus in multiple dialects and archaic forms of Estonian, Finnish, Karelian, Ingrian (Izhorian) and Votic languages. All in all, there are over 240,000 digitized texts of Finnic tetrametric oral poetry in the Finnish Literature Society and Estonian Literary Museum, and more archival texts and sound recordings in other Finnish, Estonian and Russian archives. (Harvilahti

1 Academy of Finland no. 333138, 308381, 322071 and 288119

2013; Sarv & Oras 2020; Kallio et al. 2017). In this preliminary work, we've focused on the Finnish SKVR corpus of 89,247 items in Karelian, Ingrian and Finnish languages, but we are currently working to add the Estonian corpus (see Sarv & Oras 2020), the unpublished (but digitized) Finnish corpus and some 19th-century literary works in Kalevala-meter.

The poems in SKVR were recorded from 1564 to 1939 and were originally edited and published in the 34 volumes of *Suomen Kansan Vanhat Runot* (SKVR) 'The Ancient Poems of Finnish People' (1908–1948 and 1997). The corpus is biased, for example, towards epic, narrative and poetically coherent texts (see e.g. Anttonen 2005; Tarkka 2013; Kalkun 2015; Tarkka et al. 2018; Timonen 2004), but it contains a wide variety of poetics and genres from epics and lyrical songs to incantations, ritual songs and lullabies (e.g. Harvilahti 2013; Kallio et al. 2017; Tarkka 2013).

Although not created for contemporary research questions, the corpus is unique in the scope of its documentation of local, historical Finnic oral traditions. Nevertheless, the sheer size of the data, the complex historical


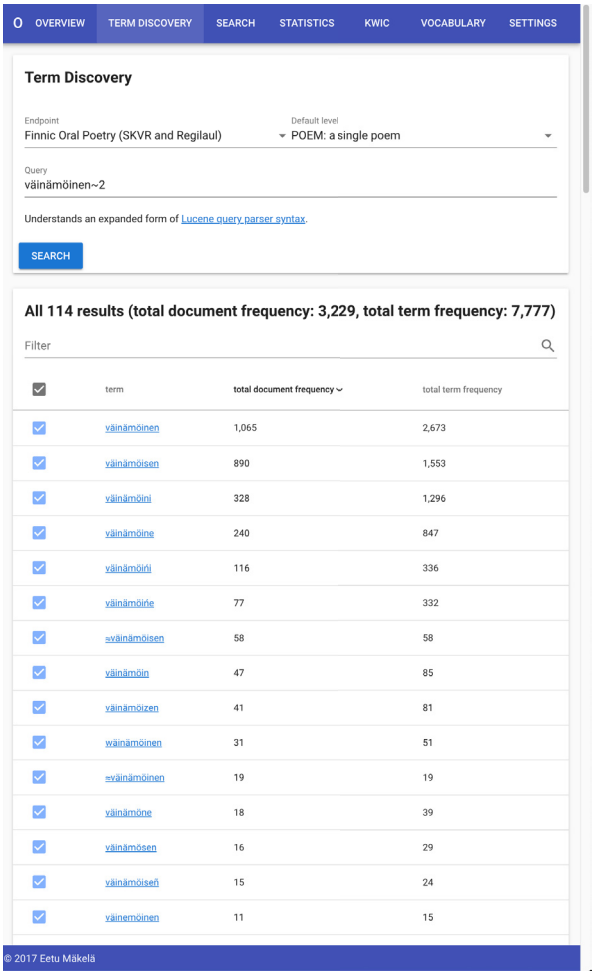


**Figure 1. First results on Octavo for one set of variations for *vanha Väinämöinen* 'Old Väinämöinen', with metadata on collector, theme ID of the type index and the place of recording, and one sentence of text around each occurrence.**

**Figure 2. Term discovery search on Octavo for one set of variations of *Väinämöinen*, the results showing how many times each variation appears in the SKVR corpus.**

and contextual knowledge needed in interpreting it, and the ample linguistic and poetic variation of texts make aims for macroscopic views (see Tangherlini 2013; 2016) difficult. The texts make use of diverse dialectal, morphological, poetic and archaic wordings, written down with various orthographies. Some folklore collectors used standard literary language, while others applied detailed phonetic transcription. Furthermore, motifs and storylines were used in versatile ways related to local understandings of poetics, genres and performance situations. (See e.g. Harvilahti 1992; Frog 2010; Timonen 2004; Saarlo 2005; Tarkka 2013; Kallio & Mäkelä 2019.) The multilevel variation and uneven quality of the data poses challenges for any computational experiments.

In the SKVR corpus, the metadata is structured, which offers possibilities for various analyses and visualisations according to the recorder of the text and the place and time of documentation. In addition, the corpus contains a typological index. Yet, the metadata also presents some problems. Although research interests today tend to concern people and society, these are not represented well in the metadata. Some dates and places of documentation are incorrect or unknown, or only vaguely identified with a region or century. The performers of the songs often remain unidentified. For the most part, the nineteenth century collectors did not think that information about informants was relevant, and many singers also preferred to remain anonymous. In the typological index, the main etic genres – like narrative poems, lyric poems, incantations, wedding songs or children's songs – have been analysed according to slightly different principles. For some genres, the index mostly reproduces those used in the printed SKVR, which in many cases were developed by the editors of the particular volumes and never unified; for others, especially lyric songs, the types are the product of recent, detailed analytical work. (See https://skvr.fi/skvr-runohakemisto.) In addition, a significant amount of essential information about the data is only found in the manuscripts, footnotes of earlier research, and prefaces of SKVR's printed volumes.

## How to Browse the Complex Corpus?

A basic need for almost any user of a corpus of texts is to be able to find individual texts – whether a particular text, comprehensive corpus or some representative examples – on the basis of some criteria, such as a certain word, formula, line, motif or poetic type, or metadata such as year, place, collector or archival signum. Small differences in the functionalities of user interfaces can thus significantly impact on what kinds of research actions are feasible. The functionalities determine the flexibility of the interface, how easy it is to move between the list of results and individual texts, how the hits are indicated and whether it possible to sort the results. In the current online SKVR database (www.skvr.fi), there are several problems for advanced use: the hits within texts are not indicated, the user cannot arrange the results by the metadata, and the possibilities for free text searches are limited (see https://skvr.fi/ohje).

In our preliminary work, the SKVR poems were loaded into the Octavo system. The Octavo system is a service Eetu Mäkelä has developed to support humanities and social science research based on combinations of large, varied and 'noisy' text corpora along with attendant metadata. The system has been developed in collaboration with multiple humanities and social science research projects. On that background, the aim has been to transcend individual datasets and questions to provide functionalities of broader relevance, while at the same time ensuring that the functionalities are able to help answer actual research questions in individual projects.

The core of the Octavo system is its rich functionalities for delineating a subset of interest out of originally large and varied datasets. These include multiple mechanisms for dealing with different types of variation in the textual content, as well as the capability to query both metadata and content at the same time. After delineating a subset of interest, the system then offers further functionalities for both close reading (as seen in Figure 1) as well as subjecting results to statistical analysis, both in terms of metadata as well as vocabulary. Further, the system has been particularly designed to support iterative workflows, where the researcher can easily experiment with and amend their query constraints in response to the results they get and the analyses they make. In addition, some result views (Figure 2) are explicitly designed to help discover new variant forms for the query terms. Due to this, a researcher can start with the most obvious and certain query forms, but through iterative improvement ensure that they are also capturing the totality of the textual phenomenon of interest, while at the same time filtering out what does not belong to it.

For the most common cases across the various humanities and social science projects, the system provides ready web-user interfaces. However, feeding these are more expressive open programmatic interfaces. Due to this, the system is able to provide its most important workflows easily for all to use, but at the same time it does not limit more tech-savvy users from amending and modifying the workflows to better suit their exact needs.

Thus far out of Octavo's functionalities, the present project has mostly used the interfaces aimed at overcoming textual variation, as well as close reading of the query results. A typical search process proceeds as a chain of different types of searches. The researcher may check the variation of some individual words (Väinä*; Väinämöinen~2) and formulas ("va* van*"~1), limit the obtained results using word forms or metadata (-vanga*; -themeID:605002230), arrange the results on the basis of metadata, take a look at only the searched verses or formulas or at longer sequences

of poems, look at the whole texts either in Octavo or the SKVR database, make similar searches on parallel verses to look for unnoticed variations of the first verse, or use the type index to find similar texts without the textual feature that has been searched for or to see how these relate to the earlier analyses. (See Kallio & Mäkelä 2019).

These kinds of search processes reveal that e.g. the name of the old sage Väinämöinen may occur in over 200 forms, including *Väinö*, *Väinämö*, *Väilämöinen*, *Viänämöinen*, *Vainämoinen*, *Wäinämöisen*, *Väinämöizen*, *Väinämyösen*, and *Väinämöinji* – of which *Väinämöinen* is the most popular with 1,017 occurrences – and with numerous inflections such as *Väinämöistä*, *Väinämöisten*, *Väinämöistennin*, *Väinämöinä* etc., sometimes added with various diacritics. In formulas and poetic lines, this kind of variation accumulates. *Väinämöinen* most often appears in the formula *vaka vanha Väinämöinen* 'steady old Väinämöinen'. Yet, he can be wise instead of steady, or the formula may get shorter to incorporate verbs or other words, such as, for example:

*Tuop oli vanha Väinämöinen*
that was old Väinämöinen

*Tuopa viisas Väinämöinen*
that wise Väinämöinen

*Olipa ennen vanha Väinö*
there once was old Väinö

*Sano vanha Väinämöinen*
said old Väinämöinen

*Päälle polven Väinämöisen*
onto the knee of Väinämöinen

The formula often has a parallel line *tietäjä iänikuinen* 'the eternal sage', which again may have inflections and variations or be replaced with other parallel formulas. Yet, if compared with some short, wide-spread sequences of formulas (*standard sequences* or *multiforms*, see Harvilahti 1992; Frog 2016), such as the ones on making a journey, the set of formulas on *Väinämöinen* is quite simple, narrow and stable (Kallio & Mäkelä 2019).

When mapping and understanding of this kind of variation is done, and various exceptions and special cases have been interpreted, the researcher has a sub-corpus to proceed with, for example, when analysing various uses of a particular formula, motif or poetic type, or the relation of these to different local or genre-specific practices, literary influences or other features.

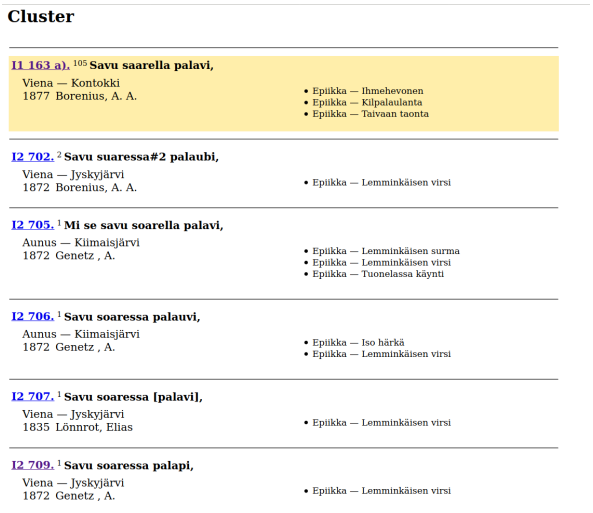


**Figure 3. Part of the cluster of the verse 'Savu soarella palaabi' ("Fire is burning on the isle").**

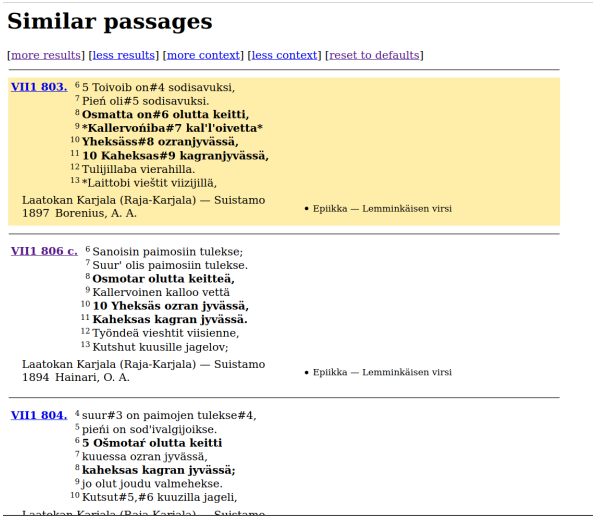


**Figure 4. Search for passages similar to "Osmatta on#6 olutta keitti, *Kallervońiba#7 kal'l'oivetta*, Yheksäss#8 ozranjyvässä, 10 Kaheksas#9 kagranjyvässä" ('Osmatta brewed beer, Kalervoini (brewed) malt-water, in nine grain of barley, in nine grain of oat').**

| | |
|---|---|
| Yht' ei kut°tsun Lemmingäistä.° | 15 Kut°tšu veri-sogeat,° |
| *Rujot (ne) reillä reissuaabi, | Ruiot re'ellä rembutteli, |
| Rammat rat°tšahin ajeli,° | Rammat rat°tšahin ajeli,° |
| | Sogeat venosin souti, |
| 20 Sogiat venozin soudi.* | Yht' ei kut°tšu Lemmingästä.° |
| Lemmingäin on poiga_piilo | 20_Lemmingäne on piilopoiga |
| Pillojah on piilemässä,#10 | Pilloja on piilemässä, |
| Pahojah pagenemassa.#11 | Pahoja pagenemassa. |
| "Hoib om moamo, kandajańi, | "Hoi on moammoni, kantajani, |
| 25_Armas maijon andajańi, | Armas maion antajani, |
| Ihalan imettäjäni, | 25_Ihala imettäjäni, |
| Et°tsib om miul pelvoi paid[a],° | Tuos miull' sot'isobani, |
| Ennembä neidona kuvottu, | Kannas paloini-paita!" |
| Kassabapeän#12 on kalkuteltu, | Emo varsin vastajeli: |
| 30 Kannabas paloni paid[a]." | Noin on#2 virkki, näin pagiši: |

**Figure 5. The side-by-side view of two automatically aligned versions of *The Song of Lemminkäinen*.**

## From Similarity of Character Bigrams to Verses, Sections and Poems

Octavo, by design, allows the user fine control in driving their discovery and exploration. However, this requires an expert user who is able and willing to put in the often significant time required to craft queries in its language, to understand its affordances and limitations and to manually keep tabs on their exploration process. Consequently, the results are still substantially dependent on the competence of the user on the variation and complications of the corpus. Thus, we are actively searching for means to make the interaction easier. Particularly, we are looking at ways to use the corpus itself to iteratively drive the search.

To this end, Maciej Janicki has started developing a prototype tool for exploring the similarity within the corpus on verse, passage and poem level. The main computational idea is to measure the similarity between individual verses as the cosine similarity on character bigrams. Roughly speaking, this amounts to how many pairs of adjacent letters the two verses have in common. For example, the verses *Armazb maijon andajańi* and *Armas maion antajani*, despite having differences in every word, have many common bigrams: "Ar", "rm", "ma" twice, "ai", "an", "aj" etc. Importantly, besides allowing for orthographic, morphological and dialectal variation, this similarity metric is also insensitive to word order and only weakly sensitive to word compounding.

After discovering the most similar pairs of verses based on bigram analysis, the verses are clustered using the Chinese Whispers algorithm (Biemann 2006), which results in groups of verses similar to each other. The Chinese Whispers algorithm starts by assigning each verse to their own group. Then, it proceeds by selecting a verse, and going through every other verse it is pairwise similar to. From the clusters that these other verses belong to, it finds the one that contains most similar verses overall to the one under evaluation, and moves the verse to that group. This is done in random order for all verses, and further repeated until no group changes occur anymore. In a network representation of the corpus, with verses being nodes and similarities between verses edges, the Chinese Whispers algorithm computes groups of nodes that are especially densely connected with each other, as compared to the rest of the network. The resulting groups of similar verses can be used to explore how a given type of verse or sequence of verses appears in the corpus regardless of surface-level variation (Figures 3 and 4).

To align two poems, the minimum edit distance algorithm (Wagner & Fischer 1974) is used. The algorithm aligns the verses between the poems in a way that maximizes the poems' overall similarity (i.e. the sum of verse-wise similarities). The same algorithm can be applied to align the paired verses themselves at the character level. The result is a side-by-side view of two poems (Figure 5), in which both the differences on the verse level (equivalent vs. non-equivalent parts) and on the character level within equivalent verses are highlighted.

The main drawback of the current approach is its inability to capture and visualize changes in verse ordering (see *Yht' ei kuttsun Lemmingäistä* in Figure 5) or to explore similarities below the verse level. Also, the bigram-based similarity metric underestimates the similarity in cases of many small phonetic differences and could be improved by taking the phonetic similarity into account (e.g. substituting a vowel with a different vowel is a much smaller difference than with a consonant). We are going to address these points in further work.

Our future idea is to test the coverage of recognising similarity by comparing the results of the interface with more manual search results on Octavo, and on the existing type index and earlier manual studies on certain poetic types. Further, it is quite essential to add possibilities for manual adjustments – what verses are most relevant, what kinds of features should count as similar or should be highlighted in comparison – and think of effective ways to visualise and interpret the similarities of large groups of verses, sections or texts. For example, Stefan Jänicke and David Joseph Wrisley (2017) visualise versions of *Chanson de Roland* in a way that helps even someone not familiar with formulaic poetry to easily understand the scope and character of variation. In short, we are experimenting with how to take the strong points of each tool and combine them into something that is both powerful as well as easier to use.

## Collaboration in Practice

Currently, research in computational social science and digital humanities rarely permeates back into their core disciplines. The problem is that current tools and approaches are often borrowed from fields where both data and research protocols are much more standardized. In the humanities, on the other hand, available datasets often have not been created for today's research, and, as a result, they are rife with complex biases. If not properly handled, these biases easily invalidate any computational research based on the corpora. Invariably, there are also gaps between what can be produced through automated means, and the nuanced human categories of interest. Thus, to produce results of interest to the subject domain, computational research by necessity would need to interleave computational inference with manual interpretation to produce the final data conclusions are based on.

Here, a challenge for a humanist is how to describe and document work processes well enough not only to give other humanists the possibility to reach similar conclusions, but to help the computational scientist to understand the process in order to make some parts of it easier. Due to the complexity of variation in the corpus, an efficient process must be equally complex and flexible, and enable the movement between quantitative views and manual interpretation of individual texts.

## Works Cited


Abello, James, Peter Broadwell & Timothy R. Tangherlini 2012. "Computational Folkloristics". *Communications of the ACM* 55(7). Pp. 60–70. <https://doi.org/10.1145/2209249.2209267>

Anttonen, Pertti 2005. *Tradition through Modernity: Postmodernism and the Nation-State in Folklore Scholarship*. Helsinki: Finnish Literature Society. <https://doi.org/10.21435/sff.15>

Arvidsson, Alf, Lauri Harvilahti, Audun Kjus, Clíona O'Carroll, Susanne Österlund-Pötzsch, Fredrik Skott & Rita Treija (eds.) 2018. *Visions and Traditions: Knowledge Production and Tradition Archives*. Helsinki: Academia Scientiarum Fennica.

Biemann, Chris 2006. "Chinese Whispers: An Efficient Graph Clustering Algorithm and Its Application to Natural Language Processing Problems". *Proceedings of TextGraphs: The First Workshop on Graph Based Methods for Natural Language Processing* (June 2006). Pp. 73–80. <https://dl.acm.org/doi/10.5555/1654758.1654774>

Frog 2010. *Baldr and Lemminkäinen: Approaching the Evolution of Mythological Narrative through the Activating Power of Expression: A Case Study in Germanic and Finno-Karelian Cultural Contact and Exchange*. UCL Eprints. London: University College London. <http://eprints.ucl.ac.uk/19428/>

Frog 2016. "Linguistic Multiforms in Kalevalaic Epic: Toward a Typology". In *The Ecology of Metre*. Ed. Ilya Sverdlov & Frog. Special issue, *RMN Newsletter* 11: 61–98.

Harvilahti, Lauri 1992. "The Production of Finnish Epic Poetry: Fixed Wholes or Creative Compositions?". *Oral Tradition* 7(1): 87–101. <http://journal.oraltradition.org/issues/7i/harvilahti>

Harvilahti, Lauri 2013. "The SKVR Database of Ancient Poems of the Finnish People in Kalevala Meter and the Semantic Kalevala". *Oral Tradition* 28(2): 223–232.

Harvilahti, Lauri 2019. "History of Computational Folkloristics in Finland and Some Current Perspectives". *Folkloristics in the Digital Age*. Ed. Pekka Hakamies & Anne Heimo. Helsinki: Academia Scientiarum Fennica. Pp. 158–175.

Hakamies, Pekka, & Anne Heimo (eds.) 2019. *Folkloristics in the Digital Age*. Helsinki: Academia Scientiarum Fennica.

Hämäläinen, Mika, Tanja Säily, Jack Rueter, Jörg Tiedemann & Eetu Mäkelä 2018. "Normalizing early English Letters to Present-Day English Spelling". *Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Ed. Alex B. Degaetano-Ortlieb et al. Stroudsburg, PA: Association for Computational Linguistics. Pp. 87–96.

Ilyefalvi, Emese 2018. "The Theoretical, Methodological and Technical Issues of Digital Folklore Databases and Computational Folkloristics". *Acta Ethnographica Hungarica* 63(1): 209–258.

Isoaho, Karoliina, Gritsenko, Daria & Mäkelä, Eetu. 2020. "Topic Modeling and Text Analysis for Qualitative Policy Research". *Policy Studies Journal*. <https://doi.org/10.1111/psj.12343>

Jänicke, Stefan, & David Joseph Wrisley 2017. "Visualizing Mouvance: Toward a Visual Analysis of Variant Medieval Text Traditions". *Digital Scholarship in the Humanities* 32(suppl_2): ii106–ii123.

Kalkun, Andreas 2015. *Seto laul eesti folkloristika ajaloos: Lisandusi representatsiooniloole*. Tartu: Eesti Kirjandusmuuseum.

Kallio, Kati, Frog & Mari Sarv 2017. "What to Call the Poetic Form: Kalevala-Meter or Kalevalaic Verse, *regivärss*, Runosong, the Finnic Tetrameter, Finnic Alliterative Verse or Something Else?' *RMN Newsletter* 12–13: 94–117. <http://hdl.handle.net/10138/305420>

Kallio, Kati, & Eetu Mäkelä 2019. "Suullisen runon sähköisestä lukemisesta". *Elore* 26(2): 25–40. <https://doi.org/10.30666/elore.84570>

Moretti, Franco 2013. *Distant Reading*. Brooklyn: Verso.

Mäkelä, Eetu, Mikko Tolonen, Jani Marjanen, Antti Kanner, Ville Vaara & Leo Lahti 2019. "Interdisciplinary Collaboration in Studying Newspaper Materiality". *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference (DHN 2019), CEUR Workshop Proceedings* 2365: 55–66. <http://ceur-ws.org/Vol-2365/07-TwinTalks-DHN2019_paper_7.pdf>

Mäkelä, Eetu, Anu Koivunen, Antti Kanner, Maciej Janicki, Auli Harju, Julius Hokkanen & Olli Seuri 2020a. "An Approach for Agile Interdisciplinary Digital Humanities Research: A Case Study in Journalism". *Proceedings of Twin Talks at the Digital Humanities in the Nordic Countries 2020. CEUR Workshop Proceedings*. <http://ceur-ws.org/Vol-2717/paper01.pdf>

Mäkelä, Eetu, Krista Lagus, Leo Lahti, Tanja Säily, Mikko Tolonen, Mika Hämäläinen, Samuli Kaislaniemi & Terttu Nevalainen 2020b. "Wrangling with Non-Standard Data". *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020), CEUR Workshop Proceedings*. <http://ceur-ws.org/Vol-2612/paper6.pdf>

Saarlo, Liina 2005. *Eesti regilaulude stereotüüpiast: Teooria, meetod ja tähendus*. Tartu: Tartu Ülikooli Kirjastus.

Säily, Tanja, Eetu Mäkelä & Mika Hämäläinen 2018. "Explorations into the Social Contexts of Neologism Use in Early English Correspondence". *Pragmatics & Cognition* 25(1): 30–49.

Sarv, Mari 2019. "Poetic Metre as a Function of Language: Linguistic Grounds for Metrical Variation in Estonian Runosongs". *Studia Metrica et Poetica* 6(2): 102–148. <https://doi.org/10.12697/smp.2019.6.2.04>

Sarv, Mari, & Janika Oras 2020. "From Tradition to Data: The Case of Estonian Runosong". *Arv: Nordic Yearbook of Folklore* 76: 105-117.

Tangherlini, Timothy R. 2016. "Big Folklore: A Special Issue on Computational Folkloristics". *Journal of American Folklore* 129(511): 5–13. <https://www.jstor.org/stable/10.5406/jamerfolk.129.511.0005>

Tangherlini, Timothy R. 2013. "The Folklore Macroscope: Challenges for a Computational Folkloristics". *Western Folklore* 72(1): 7–27. <https://www.jstor.org/stable/24550905>

Tarkka, Lotte 2013. *Songs of the Border People: Genre, Reflexivity, and Performance in Karelian Oral Poetry*. Helsinki: Academia Scientiarum Fennica.

Tarkka, Lotte, Eila Stepanova & Heidi Haapoja-Mäkelä 2018. "The *Kalevala*'s Languages: Receptions, Myths, and Ideologies". *Journal of Finnish Studies* 21(1–2): 15–45. <http://hdl.handle.net/10138/301432>

Timonen, Senni 2004. *Minä, tila, tunne: Näkökulmia kalevalamittaiseen kansanlyriikkaan*. Helsinki: SKS.

Wagner, Robert A. and Michael J. Fischer 1974. 'The String-to-String Correction Problem". *Journal of the ACM* 21(I): 168–173. <https://doi.org/10.1145/321796.321811>